



(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
31.10.2007 Bulletin 2007/44

(51) Int Cl.:
G10L 11/00 (2006.01) H04N 5/91 (2006.01)

(21) Application number: **06020505.1**

(22) Date of filing: **29.09.2006**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR
Designated Extension States:
AL BA HR MK YU

(30) Priority: **25.04.2006 US 411016**

(71) Applicant: **CyberLink Corp.**
Shindian City,
Taipei (TW)

(72) Inventors:
• **Chen, Ming-Jun**
Tai-Nan City (TW)
• **Chen, Jiun-Fu**
Changhua County 5078 (TW)
• **Tang, Shih-Min**
Tainan County 722 (TW)
• **Huang, Ho-Chao**
Shindian City
Taipei 231 (TW)

(74) Representative: **Rehberg Hüppe + Partner**
Nikolausberger Weg 62
37073 Göttingen (DE)

(54) **Systems and methods for classifying sports video**

(57) Disclosed are systems, methods, and computer readable media having programs for classifying sports video. In one embodiment, a method includes: extracting, from an audio stream of a video clip, a plurality of key audio components contained therein; and classifying, using at least one of the plurality of key audio components, a sport type contained in the video clip. In one embodiment, a computer readable medium having a computer program for classifying sports video includes: logic configured to extract a plurality of key audio components from a video clip; and logic configured to classify a sport type corresponding to the video clip.

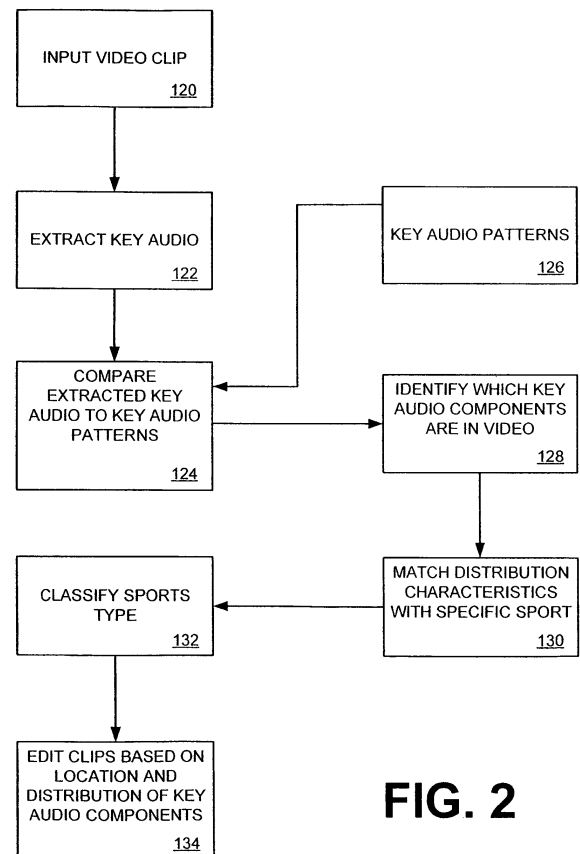


FIG. 2

Description

TECHNICAL FIELD

[0001] The present disclosure is generally related to video signal processing and, more particularly, is related to systems, methods, and computer readable media having programs for classifying sports video.

BACKGROUND

[0002] In recent years, among the various kinds of multimedia, video is becoming an important component. Video refers to moving images together with sound and can be transmitted, received, and stored in a variety of techniques and formats. Video can include many different genres including, but not limited to episodic programming, movies, music, and sports, among others. End users, editors, viewers, and subscribers may wish to view only selected types of content within each genre. For example, a sports viewer may have great interest in identifying specific types of sporting events within a video stream or clip. Previous methods for classifying sports video have required the analysis of video segments and corresponding motion information. These methods, however, require significant processing resources that may be costly and cumbersome to employ.

SUMMARY

[0003] Embodiments of the present disclosure provide a system, method, and computer readable medium having a program for classifying sports video. In one embodiment a system includes: logic configured to collect a plurality of key audio samples from a plurality of types of sports; logic configured to extract a plurality of sample audio features from a plurality of frames within each of the plurality of key audio samples; logic configured to generate a plurality of patterns corresponding to the plurality of key audio samples; logic configured to extract a plurality of audio features from the plurality of frames within an audio stream of a video clip; logic configured to compare the plurality of sample audio features in the plurality of patterns with the a plurality of audio features extracted from the audio stream; and logic configured to classify the video clip based on the location and the frequency of the key audio components.

[0004] In another embodiment, a method includes: extracting, from an audio stream of a video clip, a plurality of key audio components contained therein; and classifying, using at least one of the plurality of key audio components, a sport type contained in the video clip.

[0005] In a further embodiment, a computer readable medium having a computer program for classifying sports video includes: logic configured to extract a plurality of key audio components from a video clip; and logic configured to classify a sport type corresponding to the video clip.

[0006] Other systems and methods will be or become apparent from the following drawings and detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present disclosure. In the drawings, like reference numerals designate corresponding parts throughout the several views.

[0008] FIG. 1 is a block diagram illustrating an embodiment of building patterns for use in classifying sports video.

15 [0009] FIG. 2 is a block diagram illustrating an embodiment that uses the patterns of FIG. 1 to classify sports video.

[0010] FIG. 3 is a table illustrating exemplary embodiments of sports types as related to key audio components.

20 [0011] FIG's. 4A-4D are diagrams illustrating audio component sample strings corresponding to different sports types.

[0012] FIG. 5 is a block diagram illustrating an embodiment of a system for classifying sports video.

25 [0013] FIG. 6 is a block diagram illustrating an embodiment of a method for classifying sports video.

[0014] FIG. 7 is a block diagram illustrating an embodiment of a computer readable medium having a program for classifying sports video.

DETAILED DESCRIPTION

[0015] Reference will now be made to the drawings. While the disclosure will be provided in connection with these drawings, there is no intent to limit the disclosure to the embodiment or embodiments disclosed herein. On the contrary, the intent is to cover all alternatives, modifications, and equivalents.

35 [0016] Beginning with FIG. 1, illustrated is a block diagram of an embodiment for building patterns for use in classifying sports video. The patterns can include patterns of one or more data features for key audio components. The patterns can be compared to the same data features of video clips. In building the patterns, key audio components are collected for different sports in block 102. The key audio components can be collected for any number of sports including, but not limited to, auto racing, sumo wrestling, billiards, and figure skating, among others. Examples of key audio components contained within these types of sports can include, for example, an engine sound, a referee sound, any number of ball related collisions and a whistle, among others.

40 [0017] In block 104, features are extracted for the key audio components. Features can include, but are not limited to, mel-frequency cepstrum coefficients 106, noise frame ratio 107, and pitch 108. For example, other features that can be used include LPC coefficients 109, LSP

coefficients 111, audio energy 113, and zero-crossing rate 114. The mel-frequency cepstrum coefficients are derived from the known variation of critical bandwidths of the human ear. Filters are spaced linearly at low frequencies and logarithmically at high frequencies and a compact representation of an audio feature can be produced using coefficients corresponding to each of the bandwidths. These features can be extracted in a frame by frame manner. After the features are extracted in block 104, a pattern is built for each key audio component 110. The pattern can include the specific mel-frequency cepstrum coefficients 106 and pitch 108 that are exclusive to the key audio component. A model for each key audio component is trained in block 112 in order to capture the unique audio characteristics of the key audio component without being constrained by the limitations of a particular sample of the key audio component.

[0018] Reference is now made to FIG. 2, which is a functional block diagram illustrating use of the patterns of FIG. 1 to classify sports video. A video clip is input in block 120 and the key audio components are extracted from the video clip in block 122. The video clip can be a digital or analog streaming video signal or a video stored on a variety of storage media types. For example, the video can be stored in solid state hardware or on magnetic or optical storage media using analog or digital technology. The extracted key audio components are compared to key audio patterns 126, in block 124. The key audio components contained in the video are identified in block 128. Identifying the key audio components in the video can serve to significantly narrow the number of possible sport types and, in some cases, completely determine the sports type. For example, an engine noise will only occur in a small number of sports types, namely those performed using motorized vehicles.

[0019] Distribution and frequency characteristics corresponding to a specific sport are matched in block 130 and the video is classified as a particular sports type in block 132. The distribution and frequency characteristics in combination with the identity of the key audio components can be used to specifically classify the sports type. For example, the distribution and frequency of a ball bouncing in a basketball game are distinctive from those occurring in a tennis match. Further, optionally, the video clips can be edited based on the location and distribution of key audio components in block 134. For example, in a football game the time between plays can be edited out of a video clip by retaining the portion of the video segment that occurs starting a few seconds before a helmet collision key audio component and ending a few seconds after a whistle key audio component. One example of an audio distribution coefficient is the tempo of the occurrence of a key audio component, as expressed in, for example, the time domain.

[0020] Reference is now made to FIG. 3, which is a table illustrating exemplary embodiments of sports types as related to key audio components. The table includes a column for sports type 150, which features an example

of a variety of different sports that can be classified under the methods and systems herein. The table also includes a key audio component column 152 that lists the one or more key audio components corresponding to the sports type of column 150. The table also includes a frequency column 154, which includes an estimated frequency of occurrence corresponding to the key audio component 152 in a sports type listed in column 150. Similarly, a table also includes a distribution column 156 that categorizes a likely distribution of the key audio components within a typical event corresponding to the sports type in column 150.

[0021] By way of example, a key audio component 152 may be extracted from a sports video and identified as a whistle. In the absence of any other key audio components 152, if the whistle occurs at a low frequency throughout the event and at irregular intervals, the event may be classified based on the timing and/or regularity of the whistle (e.g. classified as a soccer match). Alternatively, where the video includes a second key audio component 152 such as a ball/floor collision or a stick/puck collision, the sport type 150 can be categorized with greater certainty (e.g. as a basketball or hockey game, respectively). Other sport types 150 may include the same key audio component 152 and only be distinguishable by frequency 154 for distribution 156. For example, the key audio component of a water splash as found in swimming and diving events may differ only in the frequency of the water splash. Alternatively, the swimming event may include a second key audio component 152 such as a starter pistol. In this manner, many different formats of the same sport can be classified through the use of the key audio components 150.

[0022] Reference is now made to FIG's. 4A-4D, which are diagrams illustrating audio component sample strings corresponding to different types of sports. Reference is first made to FIG. 4A, which is an audio component sample stream corresponding to a basketball game. Each of the bars represent an audio sample that occurs along a timeline 176. The relevance of the bars to the classification of the sports video is illustrated by the different heights of the bars. For example, a tallest bar represents a first audio component 170. A mid-sized bar represents a second audio component 172 and a short bar represents other audio 174. The first audio component 170 in this example may be a whistle. The second audio component 172 of this example might be a ball/floor collision, which occurs as a basketball being bounced on a basketball court. Note that the low-frequency of occurrence of the first audio component 170, or whistle, that occurs in an irregular distribution throughout the time of the event can occur in a basketball game and a hockey game. However, the high-frequency of occurrence at relatively regular intervals of the second audio component 172, or ball/floor collision, provides a unique audio pattern corresponding to a basketball game.

[0023] Similarly, the audio component sample string of FIG. 4B is identified as a bowling match where a first

key audio component 170 of a ball/pin collision is contained in the video at a medium frequency and with a regular distribution. FIG's. 4C and 4D both feature as first audio components 170, a water splash. However, in the diving competition of FIG. 4C, the first audio component 170 occurs relatively infrequently at a regular distribution throughout the event. In contrast, the first audio component 170 of the swimming contest has a high frequency of occurrence and a regular distribution throughout the swimming contest. Additionally, other key audio components, such as, for example, a starter pistol report, might be identified for these types of matches to further aid in classifying the sports video and optionally editing the sports video.

[0024] Reference is now made to FIG. 5, which is a block diagram illustrating an embodiment of a system for classifying sports video. The system 180 includes logic to collect key audio samples in block 182. The system 180 further includes logic to extract audio features from the key audio samples on a frame by frame basis in block 184. The audio features can be represented, for example, by audio coefficients selected from a variety of those known to one of ordinary skill in the art including, but not limited to, mel-frequency cepstrum coefficients, pitch, LPC coefficients, LSP coefficients, audio energy, zero-crossing-rate, and noise frame ratios, among others. By way of example, the mel-frequency cepstrum coefficients provide a compact representation of an audio feature that can be produced using coefficients corresponding to a specific series of bandwidths. The system 180 can utilize one or more types of features alone or in combination to define specific key audio component features that can be utilized by the logic to generate patterns in block 188.

[0025] The system 180 further includes logic to extract audio features from an audio stream of a video clip in block 190. A video clip can be a digital or analog streaming video signal or a video stored on a variety of storage media types. The audio features from the patterns are compared to the audio features from the audio stream in the logic of block 194. Based on the outcome of the comparison and the logic of block 194, the video is classified by sports type in the logic block 196. In addition to comparing the type of audio components in the patterns and the video clip, the distribution and frequency of occurrence of the key audio components can also be compared.

[0026] Reference is now made to FIG. 6, which is a block diagram illustrating an embodiment of a method for classifying sports video. The method 200 begins with extracting key audio components from a video in block 220. The key audio components contained in the video can be identified by comparison to, for example, patterns for key audio components. In this regard, patterns can be stored, for example, in a database or other data storage application that resides in any one of a variety of computer storage or processing environments or devices.

[0027] The key audio components correspond to spe-

cific identifiable audio events that occur within the audio stream of specific sports videos. For example, a key audio component of an auto racing sporting event can be an engine sound. Similarly, a key audio component of a sumo wrestling event can be the unique referee sound made at the beginning of each match. Further, the sound of a whistle may be a key audio component in a basketball game. While the whistle is not exclusive to the game of basketball, the frequency of whistle sounds combined with the distribution of the whistle sounds can be utilized in combination with other key audio components to determine whether or not a sports video is a basketball game. For example, a sports video of a hockey game may include the key audio component of a whistle and a key audio component of a skate/ice engagement sound. Alternatively, a soccer match may include a whistle sound as a key audio component with few or no other key audio components identified. In this case, the frequency of occurrence in the distribution throughout the match would be used to identify the video as a soccer match.

[0028] After extracting the key audio components, the method 200 further includes classifying a sport type contained in the video in block 230. By determining the type of key audio components, the frequency of occurrence of each of the key audio components, and the distribution of the key audio components within the video, the video can be classified as a specific sport type.

[0029] Reference is now made to FIG. 7, which is a block diagram illustrating an embodiment of a computer-readable medium having a program for classifying sports video. The computer-readable medium 300 includes logic to extract key audio component data from a video in block 320. The video can be a digital or analog streaming video signal or a video stored on a variety of storage media types. The key audio component data is extracted from an audio stream of the video.

[0030] The computer-readable medium 300 also includes logic to classify a sports type of the video in block 330. Classification is performed by comparing the key audio components from samples with the key audio component data from the video. In some embodiments a pattern corresponding to each key audio component can be used for the comparison. In this manner, the key audio components can be identified. Additionally, the frequency of occurrence within the video and the distribution throughout the video are determined for classifying the sports type of video. The distribution and frequency information can be further utilized to edit the video.

[0031] Embodiments of the present disclosure can be implemented in hardware, software, firmware, or a combination thereof. Some embodiments can be implemented in software or firmware that is stored in a memory and that is executed by a suitable instruction execution system. If implemented in hardware, an alternative embodiment can be implemented with any or a combination of the following technologies, which are all well known in the art: a discrete logic circuit(s) having logic gates for

implementing logic functions upon data signals, an application specific integrated circuit (ASIC) having appropriate combinational logic gates, a programmable gate array(s) (PGA), a field programmable gate array (FPGA), *etc.*

[0032] Any process descriptions or blocks in flow charts should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process, and alternate implementations are included within the scope of an embodiment of the present disclosure in which functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those reasonably skilled in the art of the present disclosure.

[0033] A program according to this disclosure that comprises an ordered listing of executable instructions for implementing logical functions, can be embodied in any computer-readable medium for use by or in connection with an instruction execution system, apparatus, or device, such as a computer-based system, processor-containing system, or other system that can fetch the instructions from the instruction execution system, apparatus, or device and execute the instructions. In the context of this document, a "computer-readable medium" can be any means that can contain, store, communicate, propagate, or transport the program for use by or in connection with the instruction execution system, apparatus, or device. The computer readable medium can be, for example but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, device, or propagation medium. More specific examples (a nonexhaustive list) of the computer-readable medium would include the following: an electrical connection (electronic) having one or more wires, a portable computer diskette (magnetic), a random access memory (RAM) (electronic), a read-only memory (ROM) (electronic), an erasable programmable read-only memory (EPROM or Flash memory) (electronic), an optical fiber (optical), and a portable compact disc read-only memory (CDROM) (optical). In addition, the scope of the present disclosure includes embodying the functionality of the illustrated embodiments of the present disclosure in logic embodied in hardware or software-configured mediums.

[0034] It should be emphasized that the above-described embodiments of the present disclosure, particularly, any illustrated embodiments, are merely possible examples of implementations. Many variations and modifications may be made to the above-described embodiment(s) of the disclosure without departing substantially from the spirit and principles of the disclosure.

Claims

1. A system for classifying sports video, comprising:

5 logic configured to collect a plurality of key audio samples from a plurality of types of sports;
 logic configured to extract a plurality of sample audio features from a plurality of frames within each of the plurality of key audio samples;
 10 logic configured to generate a plurality of patterns corresponding to the plurality of key audio samples;
 logic configured to extract a plurality of audio features from the plurality of frames within an audio stream of a video clip;
 15 logic configured to compare the plurality of sample audio features in the plurality of patterns with the a plurality of audio features extracted from the audio stream; and
 20 logic configured to classify the video clip based on the location and the frequency of the key audio components.

2. The system of claim 1, further comprising means for editing the video clip by using the location and distribution of the plurality of key audio components within the video clip.

3. A method for classifying sports video, comprising:

30 extracting, from an audio stream of a video clip, a plurality of key audio components contained therein; and
 classifying, using at least one of the plurality of key audio components, a sport type contained in the video clip.

4. The method of claim 3, further comprising generating a plurality of patterns corresponding to a plurality of key audio components.

5. The method of claim 4, **characterized in that** the generating further comprises collecting the plurality of key audio components that correspond to a plurality of sports and/or determining a plurality of features for each of the plurality of key audio components and/or determining a plurality of patterns of the plurality of features, wherein each pattern of the plurality of patterns corresponds to one of the plurality of key audio components and/or training a model for each of the plurality of key audio components using the corresponding pattern of features.

6. The method of one of claims 3 to 5, **characterized in that** the extracting comprises determining a plurality of features from an audio

stream of a video clip, wherein each of the features corresponds to one of a plurality of frames in the video clip and/or
 comparing the plurality of features in the audio stream with the plurality of patterns.

5

7. The method of claim 6, **characterized in that** the comparing comprises
 determining which of the plurality of key audio components is present in the video clip and/or
 determining a distribution of the plurality of key audio components that occur in the video clip.
8. The method of one of claims 3 to 7, **characterized in that** an event that generates one of the plurality of key audio components is ball strikes and/or whistles and/or car engines and/or water splashes and/or a referee sound and/or a commentator sound and/or a bell ringing.
9. A computer readable medium having a computer program for classifying ports video, comprising:
 logic configured to extract a plurality of key audio components from a video clip; and
 logic configured to classify a sport type corresponding to the video clip.
10. The computer readable medium of claim 9, further comprising logic configured to generate a plurality of key audio patterns corresponding to a plurality of key audio sample components.
11. The computer readable medium of claim 9 or 10, further comprising logic configured to determine a plurality of audio characteristics from the video clip.
12. The computer readable medium of claim 10 or 11, further comprising logic configured to compare the plurality of audio characteristics within the video clip to the plurality of key audio patterns.
13. The computer readable medium of one of claims 9 to 12, **characterized in that** the classify logic is further configured to
 determine the sport type based on which of the key audio components is in the video clip and/or
 determine the sport type based on a distribution of the key audio components within the video clip.
14. The system of claim 1 or 2, the method of one of claims 3 to 8 or the computer readable medium of one of claims 9 to 13, **characterized in that** the audio features are mel-frequency cepstrum coefficients and/or pitch and/or LPC coefficients and/or LSP coefficients and/or audio energy and/or zero-crossing-rate and/or noise frame ratio.

10

15

20

25

30

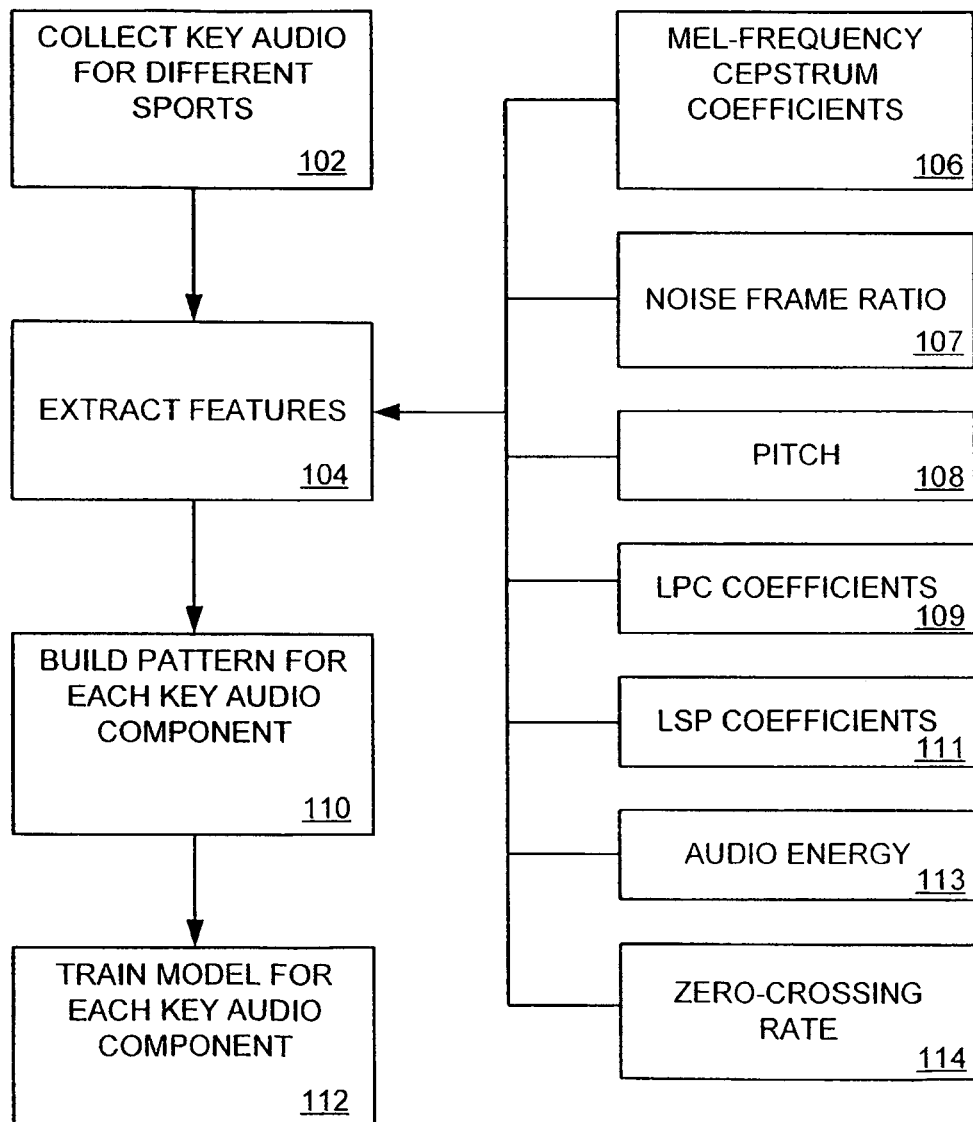
35

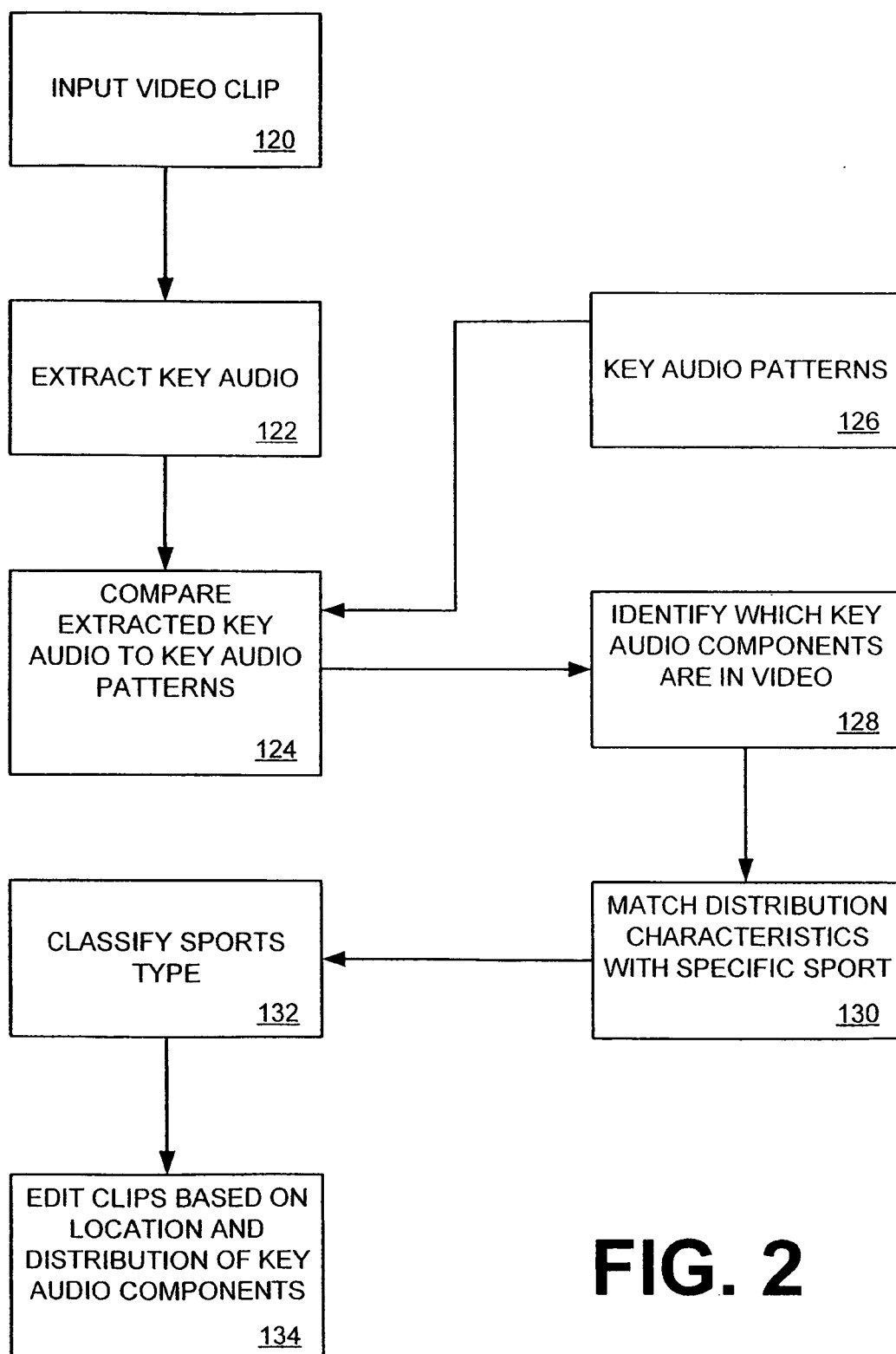
40

45

50

55

**FIG. 1**

**FIG. 2**

SPORT TYPE	KEY AUDIO COMPONENT	FREQUENCY	DISTRIBUTION
AUTO RACING	ENGINE SOUND	CONSTANT	LONG DURATION
	COMMENTATOR	INTERMITTENT	SHORT DURATION
SUMO WRESTLING	REFEREE SOUND	MEDIUM	REGULAR
TENNIS	BALL/RACKET COLLISION	HIGH	GROUPED REGULAR
BADMINTON	BIRDIE/RACKET COLLISION	HIGH	GROUPED REGULAR
BILLIARDS	BALL COLLISION	MEDIUM	REGULAR
BOWLING	BALL/PIN COLLISION	MEDIUM	REGULAR
BASEBALL	BALL/BAT COLLISION	LOW	IRREGULAR
BASKETBALL	BALL/FLOOR COLLISION	HIGH	REGULAR
	WHISTLE	LOW	IRREGULAR
SOCCER	WHISTLE	LOW	IRREGULAR
HOCKEY	STICK/PUCK COLLISION	MEDIUM	IRREGULAR
	WHISTLE	LOW	IRREGULAR
	SKATE/ICE ENGAGEMENT	HIGH	REGULAR
FIGURE SKATING	SKATE/ICE ENGAGEMENT	HIGH	REGULAR
SWIMMING	WATER SPLASH	HIGH	REGULAR
DIVING	WATER SPLASH	LOW	REGULAR
FOOTBALL	WHISTLE	MEDIUM	REGULAR
	HELMET COLLISION	MEDIUM	REGULAR

FIG. 3

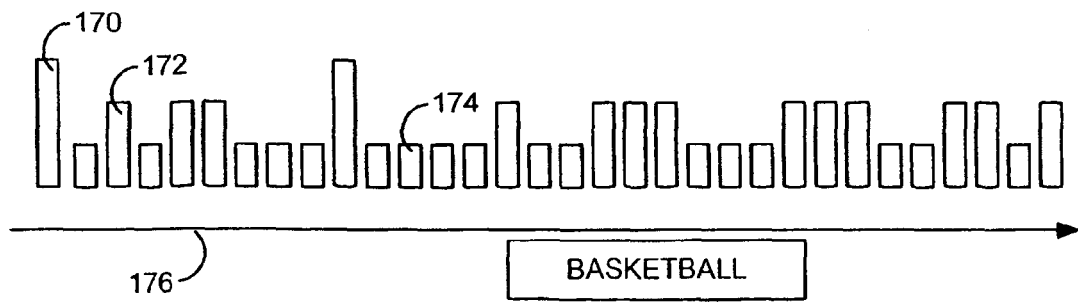


FIG. 4A

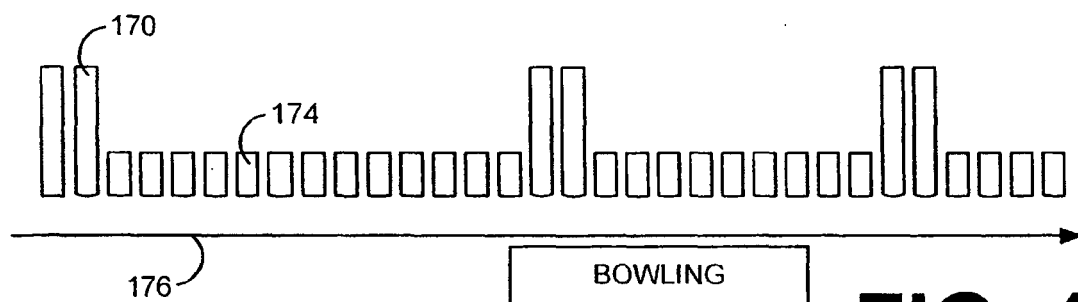


FIG. 4B

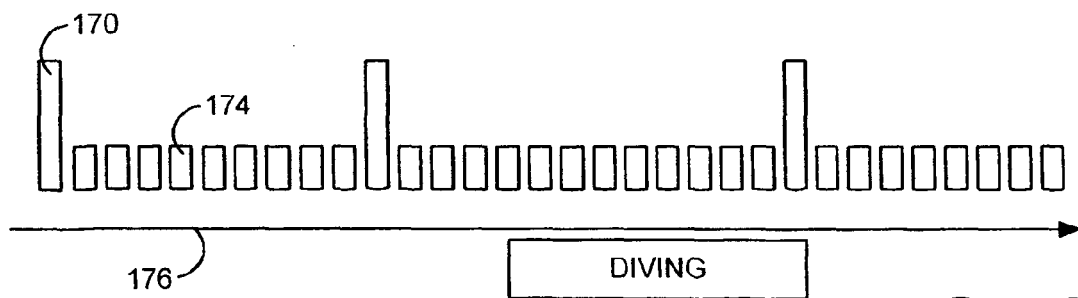


FIG. 4C

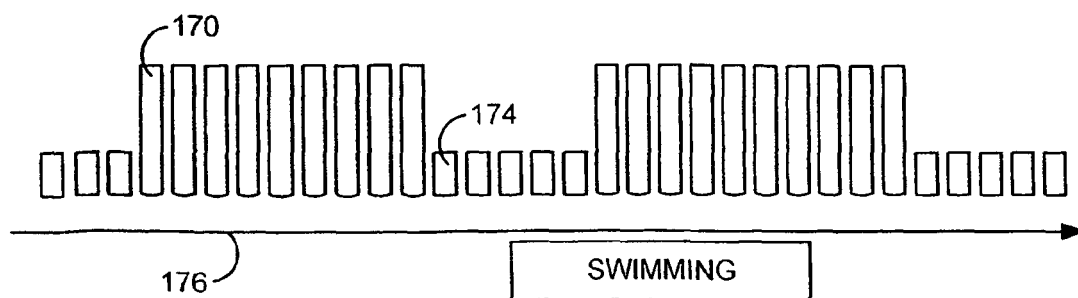


FIG. 4D

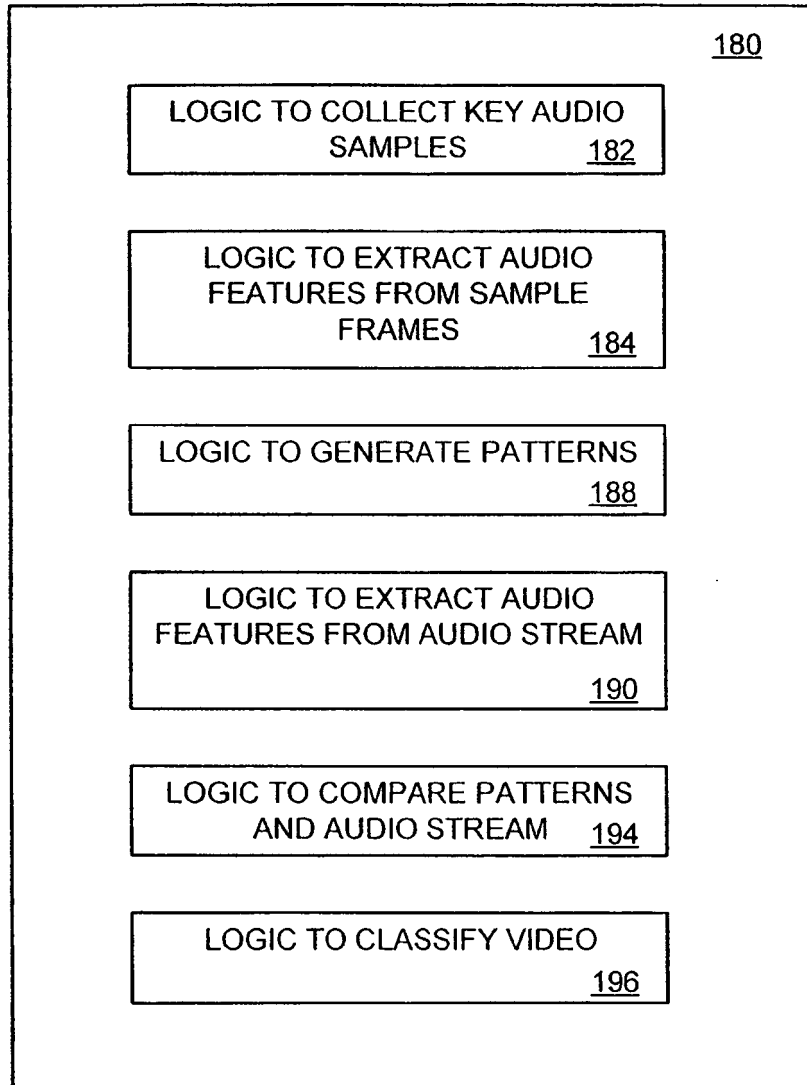


FIG. 5

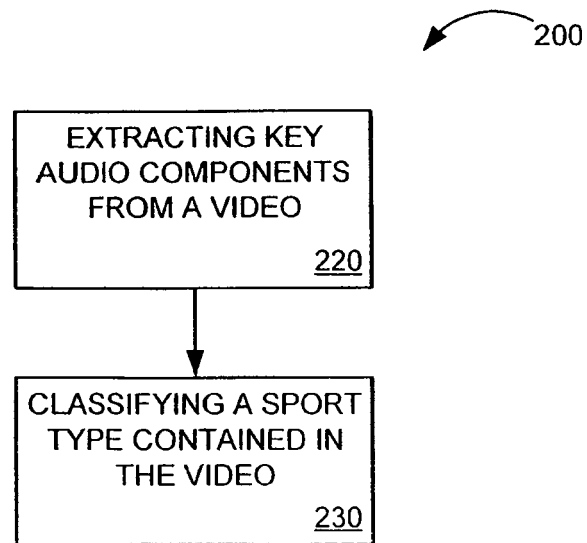


FIG. 6

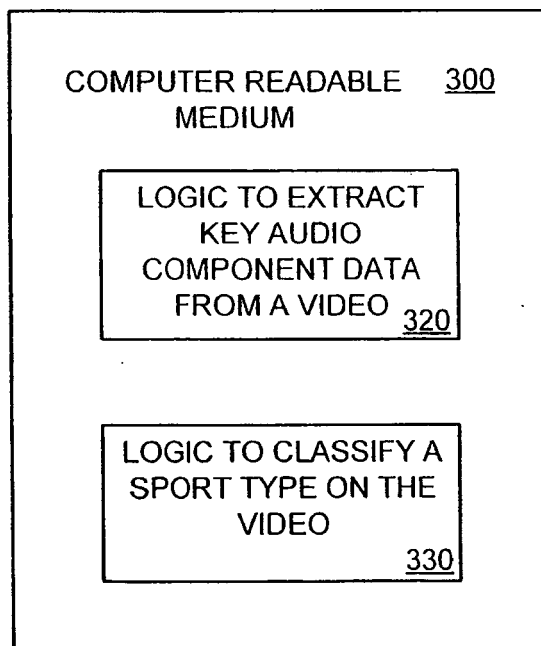


FIG. 7



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 06 02 0505

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	MESSER K ET AL: "Automatic sports classification" PATTERN RECOGNITION, 2002. PROCEEDINGS. 16TH INTERNATIONAL CONFERENCE ON QUEBEC CITY, QUE., CANADA 11-15 AUG. 2002, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, vol. 2, 11 August 2002 (2002-08-11), pages 1005-1008, XP010614050 ISBN: 0-7695-1695-X	1,3-6, 8-12,14	INV. G10L11/00 H04N5/91
Y	* the whole document *	2,7,13	
Y	EP 1 531 457 A (SONY INT EUROP GMBH [DE]) 18 May 2005 (2005-05-18) * abstract *	2,7,13	
X	ZHU LIU ET AL: "AUDIO FEATURE EXTRACTION AND ANALYSIS FOR SCENE SEGMENTATION AND CLASSIFICATION" JOURNAL OF VLSI SIGNAL PROCESSING SYSTEMS FOR SIGNAL, IMAGE, AND VIDEO TECHNOLOGY, SPRINGER, NEW YORK, NY, US, vol. 20, no. 1/2, October 1998 (1998-10), pages 61-78, XP000786728 ISSN: 0922-5773 * abstract *	1,3-6, 8-12,14	TECHNICAL FIELDS SEARCHED (IPC) G10L H04N
The present search report has been drawn up for all claims			
Place of search The Hague		Date of completion of the search 2 July 2007	Examiner Quélavoine, Régis
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

2

EPO FORM 1503 03.02 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 06 02 0505

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

02-07-2007

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1531457 A	18-05-2005	US 2005114388 A1	26-05-2005

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82